**Swin Transformer + 多任务报告**

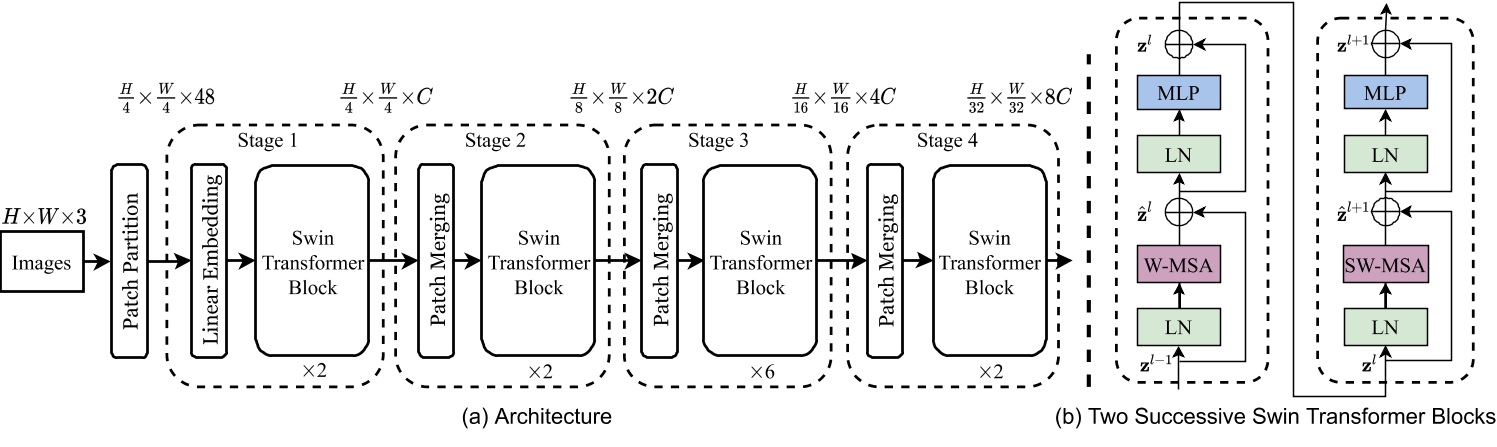
**Swin Transformer**

Transformer缺陷：

1. 图片的scale变化非常大，非标准固定
2. Transformer的自注意力计算复杂度是token数量n的平方，图片若以1像素作为一个token，计算量巨大。

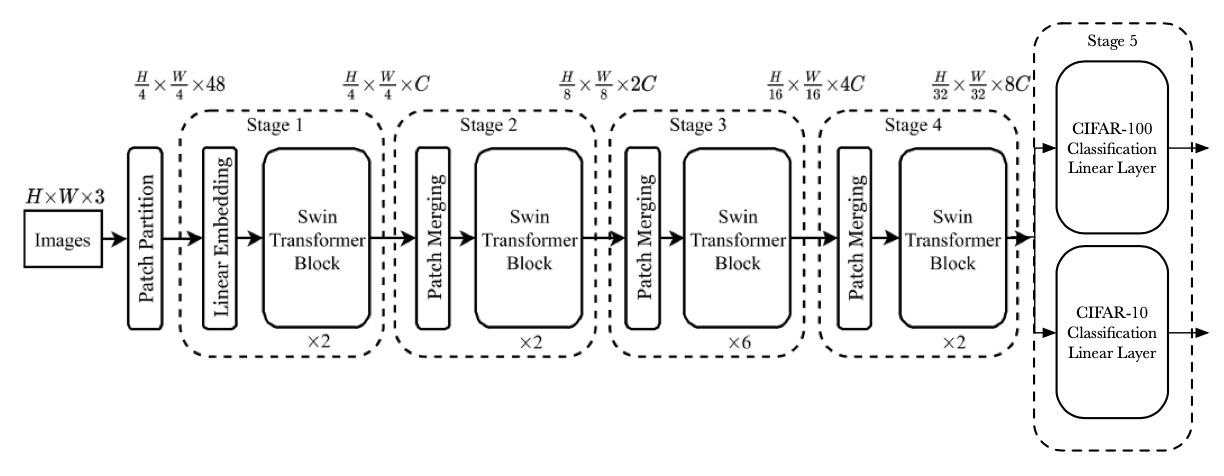
Swin Transformer的思路：

1. 通过与CNN相似的分层结构来处理图片，使得模型能够灵活处理不同scale的图片
2. 对无重合的window区域内进行self-attention计算，降低了计算复杂度。



短板：为了充分发挥Swin Transformer的性能，**需要大量的训练数据**

**Swin Transformer接多任务（CIFAR-10和CIFAR-100分类）**



多任务概述：是与单任务学习相对的一个概念。在多任务学习中，往往会将多个相关的任务放在一起来学习。一般来说有多个目标函数loss同时学习的就算多任务学习。

优势： 1、多个任务共享一个模型，占用内存量减少；

2、多个任务一次前向计算得出结果，推理速度增加；

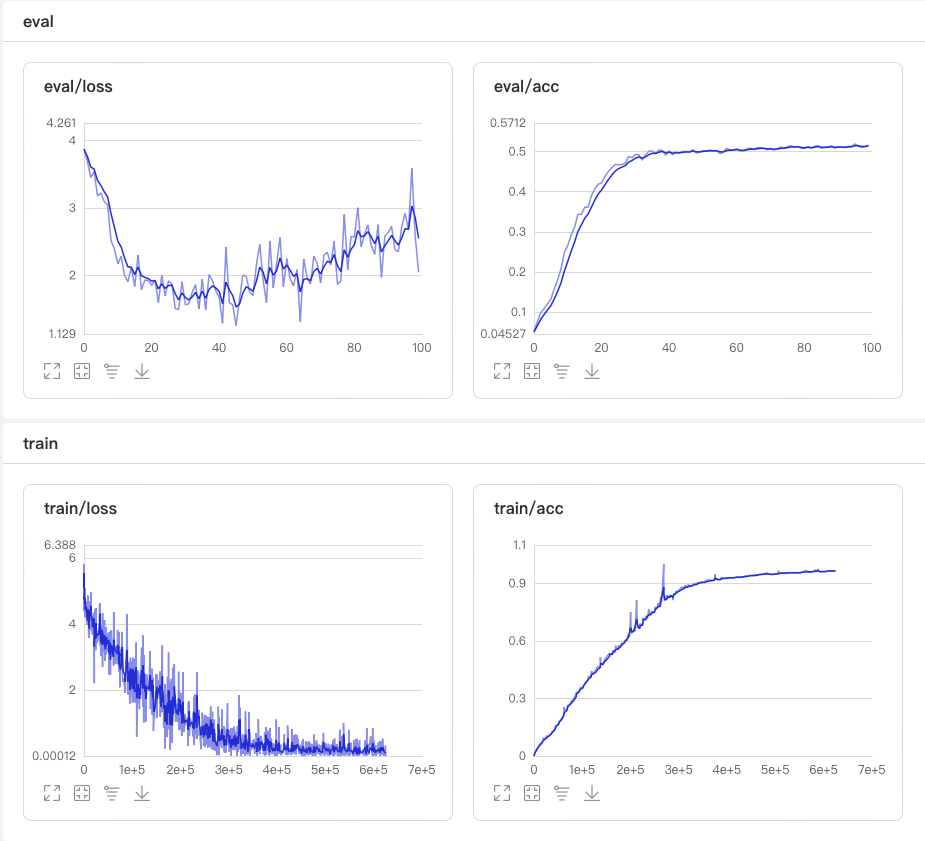
3、关联任务通过共享信息，相互补充，可以提升彼此的表现。

思路：将原用于CIFAR-100的最后一层分类头删除，为CIFAR-10和CIFAR-100分别建一个线性分类头，分别计算loss，用相加得到的总loss反向传播更新参数。

**实验**

|  |  |  |
| --- | --- | --- |
| method | CIFAR-10 acc | CIFAR-100 acc |
| Swin-T | --- | 51.9 |
| Swin-T MTL | 77.9 | 50.1 |

Swin Transformer loss和acc曲线



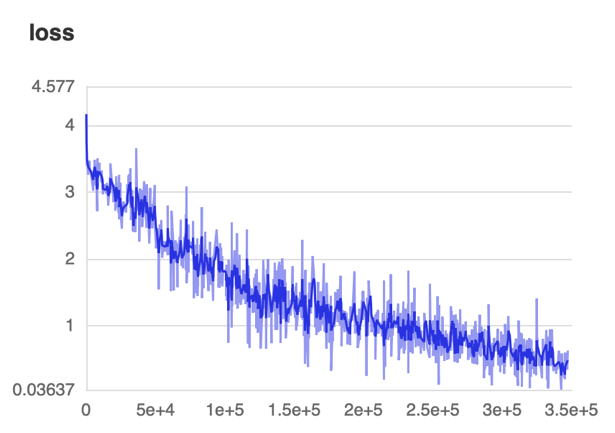
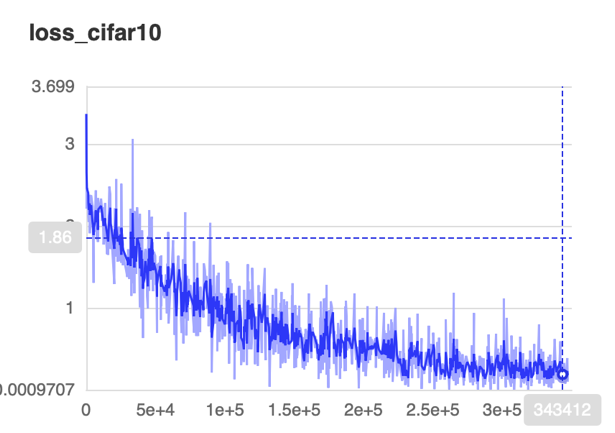
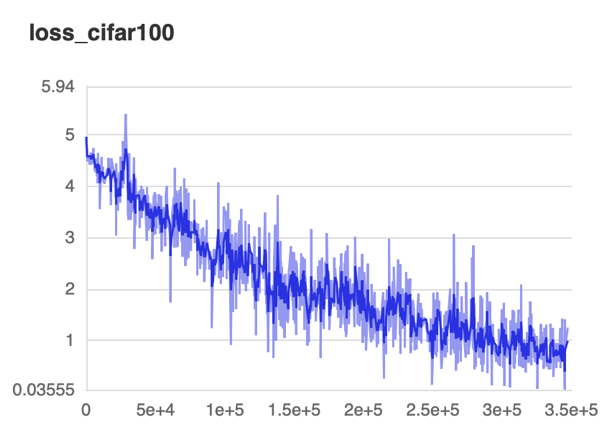
实验结果分析：

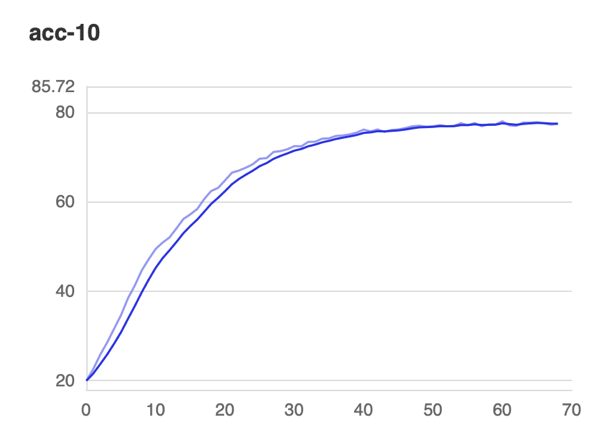
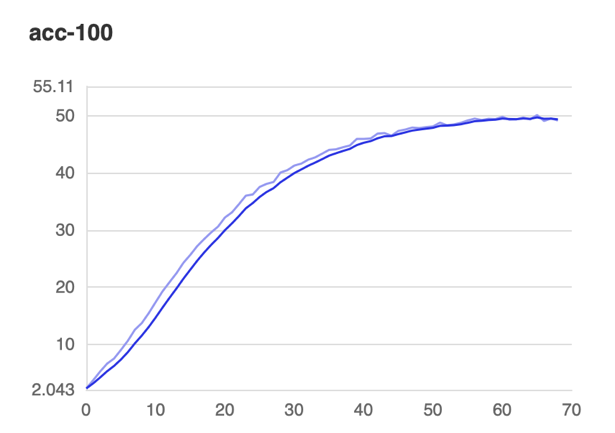
从acc分曲线可以观察到在第34个epoch左右，模型的准确率开始趋于平缓，并稳定在49%左右。

训练到第46个epoch左右时，在测试集上的损失开始增加，但准确率并没有下降。这说明模型可能出现了过拟合现象，即模型在训练集上过度拟合而在测试集上表现不佳。

在训练到90个epoch之后，训练集上的准确率超过了95%。这可能是因为模型逐渐记住了训练集中的样本，在训练集上表现出很高的准确性。如果测试集上的准确率没有持续提高或者下降了，就需要注意是否已经出现过拟合问题。

Swin Transformer MTL loss和acc曲线





实验结果分析：

该实验中，Swin Transformer模型在CIFAR-10和CIFAR-100两个数据集上进行了图像分类多任务学习，并使用两个数据集上的平均损失作为总损失来更新模型参数。

在CIFAR-10数据集上，模型的准确率在训练到第55个epoch左右趋于平稳，稳定在77%左右。在CIFAR-100数据集上，准确率在训练到第57个epoch左右趋于平稳，稳定在49%左右。

通过将CIFAR-10和CIFAR-100两个数据集上的损失取平均并作为总损失进行优化，可以使模型在不同任务上实现一定程度的共享、迁移学习。从结果来看，模型在两个数据集上都取得了相对较好的准确率，表明该多任务训练的效果不错。